DOCUMENT RESUME

ED 383 728 TM 023 158

AUTHOR Wang, Jianjun
TITLE Critical Values of Guessing on True-False and Multiple-Choice Tests.
PUB DATE Apr 95
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bayesian Statistics; *Grading; *Guessing (Tests); Models; *Multiple Choice Tests; *Objective Tests; Scores; Testing Problems; *Test Interpretation

ABSTRACT

Effects of blind guessing on the success of passing true-false and multiple-choice tests are investigated under a stochastic binomial model. Critical values of guessing are thresholds which signify when the effect of guessing is negligible. By checking a table of critical values assembled in this paper, one can make a decision with 95% confidence as to whether the correction for guessing is necessary for grading a true-false or multiple-choice test. This useful tabulation also represents an intermediate step toward development of a more comprehensive model for non-blind guessing in Bayesian statistics. (Contains 18 references and 2 tables.) (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

PERMISSION TO REPRODUCE THIS MATERIAL IN **OTHER THAN PAPER COPY** HAS BEEN GRANTED BY

Jianjun Wang

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 383 728

Critical Values of Guessing on True-False and Multiple-Choice Tests

Jianjun Wang
Department of Teacher Education
School of Education
California State University, Bakersfield

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

RUNNING HEAD: CRITICAL VALUES OF GUESSING

Paper presented at the 1995 Annual Meeting of the American Educational Research Association, April 18-22, San Francisco, California.

BEST COPY AVAILABLE

Critical Values of Guessing on True-False and Multiple-Choice Tests

Abstract

Effects of blind guessing on the success of passing true-false and multiple-choice tests are investigated under a stochastic binomial model. Critical values of guessing are thresholds which signify when the effect of guessing is negligible. By checking a table of critical values assembled in this paper, one can make a decision with 95% confidence as to whether the correction for guessing is necessary for grading a true-false or multiple-choice test. This useful tabulation also represents an intermediate step toward development of a more comprehensive model for non-blind guessing in Baysian statistics.

Critical Values of Guessing on True-False and Multiple-Choice Tests

Correction for guessing has been a persistent problem in the interpretation of true-false and multiple-choice test scores. Many authors have maintained that no solution to this problem is in sight. Thorndike (1971) pointed out: "Practice in United States testing organizations and among test publishers with respect to using the correction formula remains divided" (p.59). Payne (1992) concurred: "Researchers have for more than 30 years been investigating the problem of whether or not to correct for guessing. There is still no definite answer or agreement among the experts" (p.108).

One approach to the correction for guessing has been to investigate the conditions under which the influence of blind guessing on the scores of a test is negligible. Sax (1989) pointed out that teachers should include more items in tests to ignore the effect of guessing. Hopkins and Stanley (1981) asserted: "It should be evident that the greater the number of options per item, the less likely it is that one will select the correct option by chance and, hence, the less the magnitude of the weighting of an incorrect response" (p.149). Most researchers agreed that the influence of blind guessing on the scores of a test diminishes as the length of a test and the number of options per item increase (e.g., Ebel and Frisbie, 1991; Brown, 1981; and Mehrens and Lehmann, 1984).

Nonetheless, when the correction for guessing is ignored, it becomes possible that a student may pass a test through guessing. In terms of statistics, a null hypothesis (H_o) may be formulated that the effect of guessing is negligible. The mistake of ignoring the role of guessing when the effect of guessing does exit is called Type I error. In social sciences, the acceptable risk of making Type I error is conventionally set at $\alpha = .05$.

Critical value is a statistic that marks the edge of the rejection region of H_o at $\alpha = .05$ (Heiman, 1992). In a long true-false or multiple-choice test, the probability of obtaining a high score through guessing is small (Sax, 1989). The passing score of a test is the statistic that controls the risk of Type I error. The higher the passing score, the less the risk of rejecting the null hypothesis (H_o). The rejection region of H_o contains scores at which the probability of passing a test through guessing is less than 5%. The lowest passing score which guarantees a no larger than 5% risk is the critical value of a passing score for correction of guessing. By checking whether a passing score of a test is higher than the corresponding critical value, a decision can be made with 95% confidence as to whether the correction for guessing is necessary. Accordingly, although no solution to the correction for guessing is in sight, it is possible to construct a table of critical values to evaluate the effect of guessing.

Purpose

The critical value of a passing score is determined by the structure of a test and the stochastic model of guessing in which the probability of passing the test through guessing is delineated. Nevertheless, no such stochastic model has been stressed in educational and psychological measurements yet, needless to mention the construction of critical values to meet the structure of various tests (Brown, 1981; Mehrens & Lehmann, 1987). The purpose of this paper is to build a stochastic model describing the probability of passing a true-false or multiple-choice test through guessing, and to assemble a table of critical values for commonly used standardized tests.

The application of the table is straightforward. For a test with given total number of items (N) and the probability of guessing an item correctly (p), the table contains a critical value (x_o) identified from the stochastic model. Based on the rationale of hypothesis testing, the correction for guessing can not be ignored unless the passing score (x) of the test has been set at an $x > x_o$ level. Hence, the critical value x_o acts as a threshold which determines when the effect of guessing is negligible.

Stochastic Model

The probability of successfully passing a test through blind guessing can be modeled as a coin-tossing process. The head and

tail from the coin-tossing are two events corresponding to the success and failure in an item-guessing process respectively. Given that an item has n options, the probability of obtaining the correct option through blind guessing is 1/n. Because the number of options in each item is no less than 2, the probability of guessing an answer correctly is no larger than 50% in general. Thus, the tossed coins are unbalanced. The relations between the number of options in an item and the probability of guessing the answer correctly are listed in Table 1.

Table 1 inserted around here

The coin-tossing process is an elementary stochastic process, and has been readily solved in most math-statistical textbooks (Casella & Berger, 1990). Since most commonly used standardized tests have the same number of options per item, the probability of obtaining a correct answer through guessing does not change from item to item. For a test which contains N different items, the results of taking the test through blind guessing are equivalent to the events of independently tossing a set of N identical coins.

In statistics, a single coin tossing is a Bernoulli trial, and the entirety of tossing N coins follows a binomial (N,p) distribution with p equal to the probability of head in each

trial (Bhat, 1984). Thus, the probability of having x heads in N trials is:

$$P(X=x|N,p) = \binom{N}{x} p^x (1-p)^{(N-x)} \qquad x=0,\ 1,\ 2,\ \ldots,\ N \qquad (1)$$

The total number of items (N) and the number of options per item (n) are structural characteristics of a test. The probability of guessing an item correctly (p) is determined in Table 1 by the number of options per item (n). The events of passing a test include cases in which one obtains a score higher than a passing score. Critical value is the lowest passing score at which the probability of passing the test through guessing is less than .05. Hence, the cumulative probability of obtaining a score higher than the critical value through guessing can be calculated using Binomial (N, p) distribution. For a given N and p, the critical value (x_0) follows formula (2):

$$\Sigma_{x>x_0} \binom{N}{x} p^x (1-p)^{(N-x)} = \alpha \qquad (2)$$

The construction of critical values based on formula (2) needs the cumulative sums of terms of the binomial distribution. Eisenhart (1949) pointed out:

> Cumulative sums of terms of the binomial distribution can be obtained directly from Tables of the Incomplete Beta-Function (Edited by Karl Pearson, Biometrika Office, University College, London, 1934), but owing to the conflict between the notation of the tables and that commonly used

> for the binomial distribution, the extraction of a binomial probability from the tables is particularly difficult on each new occasion, and even for continual use requires patience and care.(p. IV)

Ferris (1944) accorded:

> For sample sizes up to 50, generally including the first sample size in a double-sampling plan, the required binomial values can be read directly from Karl Pearson's compilation. However, for second sample sizes above 50 and quality levels in the range stated above, no tabulations of any scope were available.(p.1)

Fortunately, the Pearson's table has been converted into *Tables of the Binomial Probability Distribution* by National Bureau of Standards (1950) for sample size (N) equal to 1, 2, ..., 49. Ballistic Research Laboratory (1944) also assembled *Tables of Binomial Probabilities* for N equal to 60, 75, 90, 100, 150, 200, 250 and 300. According to Burington and May (1970), these are the extensive tables of binomial distribution.

Non-blind guessing can be modeled in a Baysian stochastic process. Because a student may have partial knowledge, an informative guess could be made in a true-false or multiple choice test. Based on the Bayesian statistics, the chance of committing the guess can be described in a conditional probability. Given the condition that the guess has been made, the probability of making a correct guess can be simplified in a binomial distribution (Casella & Berger, 1990). Thus, tabulation of the simple binomial model represents an indispensable step toward development of the more comprehensive guessing model for non-blind guessing in Baysian statistics.

A Table of Critical Values

The critical values constructed in this paper are based on the two tables congregated by National Bureau of standards (1950) and Ballistic Research Laboratory (1944). Because the criterion of $\alpha = .05$ is set in formula (2), not every group of N and p has a critical value x_o. For example, for a test with a small N and large p, such as N=4 and p=.5, the probability of obtaining a full score through guessing is .0625, a value larger than .05. Thus, no matter what the passing score has been chosen, the effect of guessing for this test is not negligible at $\alpha = .05$ level. The same situation exits for a test with N=3 and p=.5, or N=4 and p>.22 structures.

It should be further noted that the score of a multiple-choice test, including the passing score, is an integer counted on the number of correctly answered items. However, the critical value (x_o) calculated from formula (2) may not be an integer. Fractional values of x_o are not physically interpretable because critical values represent the minimum passing scores which can not be achieved by blind guessing at $\alpha = .05$ level. To guarantee that the risk probability is no larger than α, the critical value (x_o) calculated from formula (2) is rounded up to an integer. As a result, with a level of passing score higher than the critical value, the risk of allowing a student passing a test through guessing is less than α.

Critical Values of Guessing

Critical values (x_o) of passing scores for commonly used standardized tests are listed in Table 2 with the probability of guessing an item correctly (p) identified by the number of options per item (n) and the length of a test represented by the total number of test items (N).

Table 2 inserted around here

The implication of Table 2 is two-fold. First, it has been shown that the critical value (x_o) increases as the number of options per item (n) decreases. Secondly, while critical values (x_o) increase along with the length of a test (N), the ratio x_o/N generally decreases as the N increases. Hence, it is demonstrated in Table 2 that the effect of blind guessing diminishes as the number of option per item and the length of a test increase.

Discussion

Kane (1994) pointed out: "The validity of test-based decisions about readiness for a course or a profession depends on the appropriateness of the passing scores used to make the decisions" (p.425). The critical value of passing score presented in Table 2 is an instrument to measure the effect of guessing in a limited number of true-false or multiple choice tests. For a test with length (N) and the number options per

item (n) not listed in Table 2, Formulae (3) and (4) can be employed to construct the critical value (Casella & Berger, 1990).

$$\Sigma_{k=0}^{x} \binom{N}{k} p^{k}(1-p)^{(N-k)} = (N-x)\binom{N}{x}\int_{0}^{1-p} t^{N-x-1}(1-t)^{x}dt \tag{3}$$

$$P(X=x) = \frac{(N-x+1)}{x}\frac{p}{1-p}P(X=x-1) \tag{4}$$

Formula (3) is based on an extensive table of Incomplete Beta-Distribution. Formula (4) is a recursion equation to augment the list of critical values. The reason for using (3) and (4) rather than a linear interpolation is that "linear interpolation will generally not be accurate to more than two decimal places, and sometimes less" (Burington & May, 1970, p.351).

References

Ballistic Research Laboratory. (1944). Tables of binomial probabilities, Memorandum Report No. 309. Aberdeen Proving Ground, Md., Ordnance Research Center, Project No. 4457.

Bhat, N. U. (1984). Elements of applied stochastic processes (2nd ed.). NY: Wiley.

Brown, F. G. (1981). Measuring Classroom Achievement. NY: Holt, Rinchart and Winston.

Burington, R. S. and May, D. C. (1970). Handbook of probability and statistics with tables (2nd ed.). NY: McGraw-Hill Book Company.

Casella, G and Berger, R. L. (1990). Statistical inference. Pacific Grove, CA: Brooks/Cole Publishing Company.

Ebel, R. L. and Frisbie, D. A. (1991). Essentials of educational measurement (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Eisenhart, C. (1949). Foreword. In National Bureau of Standards (Ed.), Tables of the binomial probability distribution. Washington, D. C.: United States Government Printing Office.

Ferris, C. D. (1944). Tables of binomial probabilities. In Ballistic Research Laboratory (ed.), Tables of binomial probabilities, Memorandum Report No. 309. Aberdeen Proving Ground, Md., Ordnance Research Center, Project No. 4457.

Heiman, G. W. (1992). Basic statistics for the behavioral sciences. Boston: Houghton Mifflin Company.

Hopkins, K. D. and Stanley, J. C. (1981). Educational and psychological measurement and evaluation (6th ed.). Englewood Cliffs, NJ: Prentice Hall.

Lindquist, E. F. (1951). Educational measurement. Washington, D.C.: American Council on Education.

Mehrens, W. A. and Lehmann, I. J. (1984). Measurement and evaluation in education and psychology (3rd ed.). NY: Holt, Rinehart and Winston.

Mehrens, W. A. and Lehmann, I. J. (1987). Using standardized tests in education. NY: Longman.

National Bureau of Standards. (1950). Tables of the binomial probability distribution. Washington, D. C.: United States Government Printing Office.

Payne, D. A. (1992). Measuring and evaluating educational outcomes. NY: Maxwell Macmillan International.

Pearson, K. (1934). Tables of the incomplete beta-function. Cambridge, England: The University Press.

Sax, G. (1989). Principles of educational and psychological measurement and evaluation (3rd ed.). Belmont, CA: Wadworth Publishing Company.

Thorndike, R. L. (1971). Educational measurement (2nd ed.). Washington, D.C.: American Council on Education.

TABLE 1

Probability of Guessing a Correct Answer to an n-choice Test Item

Number of Choices	2	3	4	5	6	7	8	9	10
Probability	.50	.33	.25	.20	.17	.14	.13	.11	.10

TABLE 2

Critical Value of Passing Score for an N-item, n-Choice Test

Number of items (N)	Number of Choices (n)								
	2	3	4	5	6	7	8	9	10
1	-	-	-	-	-	-	-	-	-
2	-	-	-	2	2	2	2	2	2
3	-	3	3	3	3	3	2	2	2
4	-	4	4	3	3	3	3	3	3
5	5	4	4	4	3	3	3	3	3
6	6	5	4	4	4	3	3	3	3
7	7	5	5	4	4	4	4	3	3
8	7	6	5	5	4	4	4	3	3
9	8	6	5	5	5	4	4	4	4
10	9	7	6	5	5	4	4	4	4
11	9	7	6	6	5	5	4	4	4
12	10	8	7	6	5	5	5	4	4
13	10	8	7	6	6	5	5	4	4
14	11	9	7	6	6	5	5	5	4
15	12	9	8	7	6	5	5	5	5
16	12	9	8	7	6	6	5	5	5
17	13	10	8	7	7	6	6	5	5
18	13	10	9	8	7	6	6	5	5
19	14	11	9	8	7	6	6	6	5
20	15	11	9	8	7	7	6	6	5
21	15	12	10	8	8	7	6	6	6
22	16	12	10	9	8	7	7	6	6
23	16	12	10	9	8	7	7	6	6
24	17	13	11	9	8	7	7	6	6

TABLE 2 (cont.)

Number of items (N)	Number of Choices (n)								
	2	3	4	5	6	7	8	9	10
25	18	13	11	9	8	8	7	6	6
26	18	14	11	10	9	8	7	7	6
27	19	14	12	10	9	8	8	7	6
28	19	14	12	10	9	8	8	7	7
29	20	15	12	10	9	8	8	7	7
30	20	15	13	11	10	8	8	7	7
31	21	16	13	11	10	9	8	7	7
32	22	16	13	11	10	9	8	8	7
33	22	16	13	12	10	9	9	8	7
34	23	17	14	12	11	9	9	8	7
35	23	17	14	12	11	9	9	8	8
36	24	18	14	12	11	10	9	8	8
37	24	18	15	13	11	10	9	8	8
38	25	18	15	13	11	10	10	9	8
39	26	19	15	13	12	10	10	9	8
40	26	19	16	13	12	10	10	9	8
41	27	20	16	14	12	11	10	9	8
42	27	20	16	14	12	11	10	9	9
43	28	20	17	14	13	11	10	9	9
44	28	21	17	14	13	11	11	9	9
45	29	21	17	15	13	11	11	10	9
46	30	22	17	15	13	11	11	10	9
47	30	22	18	15	13	12	11	10	9
48	31	22	18	15	14	12	11	10	9

TABLE 2 (cont.)

Number of items (N)	Number of Choices (n)								
	2	3	4	5	6	7	8	9	10
49	31	23	18	16	14	12	11	10	10
60	37	28	22	18	16	14	13	12	11
75	46	33	26	22	19	17	16	14	13
90	54	38	30	25	22	19	18	16	15
100	59	42	33	28	24	21	20	17	16
150	86	60	47	39	34	29	28	24	23
200	113	79	61	50	44	37	35	30	28
250	139	96	75	62	54	45	42	37	34
300	165	114	88	73	63	53	50	43	40

BEST COPY AVAILABLE